Introduction

Rationale

Autism spectrum disorder (ASD) is a complex developmental disorder that occurs in approximately 1 in 68 children (Centers for Disease Control and Prevention (CDC), 2014) and results in significant challenges with social skills, communication and behaviour (American Psychiatric Association, 2013). Individuals with ASD typically exhibit a number of stereotyped behaviours or interests including compulsions, echolalia and motor stereotypies such as hand flapping and body rocking (Bodfish et al., 2000) as well as difficult behaviours such as self-harm, aggression and non-compliance (Fox et al., 2002; Singh et al., 2006). These maladaptive behaviours are typically dealt with through multiple interventions of varying intensity, including speech-language therapy, occupational therapy, physical therapy and behavioural interventions (e.g. applied behaviour analysis; Myers and Johnson, 2007; Simpson et al., 2005). Many studies have demonstrated the effectiveness of these interventions at improving the outcomes for individuals with ASD, particularly if the interventions are intensive and introduced early in life (Corsello, 2005; McConachie and Diggle, 2007; Virués-Ortega, 2010). Despite the evidence supporting these programmes, early intervention can be extremely cost prohibitive for the families of children with ASD (Chasson et al., 2007), thus placing a significant financial burden on individual families and social services. Furthermore, many families experience long wait times for publicly funded services, which can...
negatively impact both the child’s level of functioning and the family’s quality of life (Brown, 2006). Thus, there is a critical need for additional evidence-based forms of treatment to help reduce the maladaptive behaviours associated with ASD and promote the positive behaviours necessary for engagement at home, school and in the community.

Previous research has demonstrated that regular engagement in physical activity has a positive impact on the psychological well-being of individuals with typical development (Fedewa and Ahn, 2011; Penedo and Dahn, 2005; Strong et al., 2005). For example, engagement in physical activity has been found to be associated with reduced anxiety and depression and improved self-concept, concentration, memory and academic performance (Fedewa and Ahn, 2011; Penedo and Dahn, 2005; Strong et al., 2005). Despite the growing evidence of the mental and behavioural benefits of physical activity for individuals with typical development, there is still limited research into the benefits for individuals, particularly children and youth, with ASD. The maladaptive behaviours associated with ASD can sometimes be attributed to stress, anxiety and other comorbid psychiatric disorders (Myers and Johnson, 2007; Simonoff et al., 2008). Since mental health can be improved through physical activity in individuals with typical development (Penedo and Dahn, 2005), there is the potential that exercise interventions could help to improve the mental health, and behavioural functioning, in individuals with ASD. Exercise interventions also have the potential to be much more cost-effective than traditional behavioural therapies, which require highly trained specialists and intensive one-on-one support. In contrast, exercise can be performed in one’s home or community, with little-to-no equipment or specialized support. It is important to better understand the state of the literature in this regard in order to create and inform evidence-based treatment plans for children with ASD.

Four previous reviews have been conducted regarding the impact of physical activity and exercise interventions on individuals with ASD (Lang et al., 2010; Petrus et al., 2008; Sorensen and Zarrett, 2014; Sowa and Meulenbroek, 2012) and all have found positive benefits; however, none of the reviews have used a comprehensive peer-reviewed search strategy, assessed internal validity with the Scottish Intercollegiate Guidelines Network (SIGN) criteria (Harbour and Miller, 2001), included comprehensive behavioural outcomes or focused solely on children and youth aged under 16 years. Given the exponential increase in the amount of research conducted on individuals with ASD, it is important to periodically review the most recent intervention research; the previously published reviews all considered research published before 2012 (Lang et al., 2010; Petrus et al., 2008; Sorensen and Zarrett, 2014; Sowa and Meulenbroek, 2012). Our review will examine the impact of exercise interventions on the behavioural outcomes of individuals aged 0–16 years with ASD; this is unique from previous reviews in a number of ways. First, both Lang et al. (2010) and Sowa and Meulenbroek (2012) examined interventions for children and adults with ASD up to 41 years of age, this reduces the generalizability of the results to any one age group. Petrus et al. (2008) examined the impact of exercise interventions on the stereotypic behaviour of individuals with ASD aged under 19 years and found reductions in stereotypic behaviour. Stereotypic behaviours are only one element of overall behaviour in children with ASD; therefore, the current review will be more comprehensive by including other elements of overall behaviour such as attention and social-emotional behaviours. The most recent review conducted by Sorensen and Zarrett (2014) focused on the benefits of physical activity for adolescents with ASD; however, much of the focus was on the medical health and fitness-related benefits, in addition to some behavioural benefits.

In contrast, this review will focus solely on the impact of exercise interventions on a range of behavioural symptoms including stereotypic and maladaptive behaviours, as well as positive social behaviours. Furthermore, this review will synthesize the literature with regard to younger children and youth, who have often been left out of previous reviews. Finally, we will use a rigorous systematic approach, including a two-stage screening process and assessment of bias, to fill these critical gaps in the literature.

**Objectives**

The purpose of this systematic review was to critically examine and synthesize the literature with regard to the impact of exercise interventions on the overall behaviour of individuals with ASD from birth to 16 years of age.

**Method**

**Registration**

This review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 22 July 2014 (CRD42014010811).

**Eligibility criteria**

**Population.** Our review focused on studies of individuals aged 0–16 years with a diagnosis of ASD (American Psychiatric Association, 2013), including diagnoses of pervasive developmental disorder (PDD) made under the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; DSM-IV-TR; American Psychiatric Association, 2000).

**Interventions.** This review was restricted to studies that examined the effectiveness of exercise interventions. Exercise was defined as ‘physical activity that is planned,
structured, repetitive, and purposeful’ (Caspersen et al., 1985: 128). This definition was chosen because of its broadness and largely inclusive nature for the many forms of exercise that can be used in a therapeutic or recreational setting for individuals with ASD. Studies were excluded in which exercises were used in an educational capacity only, for example, suggestions to engage in exercise.

**Comparison groups.** Studies with and without a comparison group were included in this review. We included studies that compared physical exercise interventions to comparison groups with a similar diagnosis and no intervention, as well as with a similar diagnosis and a different intervention.

**Outcomes.** To be eligible, studies had to include one or all of the following outcomes: (1) measure of stereotypic behaviours, (2) measure of cognition or attention, (3) measure of social-emotional behaviours and (4) measure of additional behavioural outcomes.

**Study characteristics.** Eligible studies met the following criteria: (1) English language, (2) at least one participant with a diagnosis of ASD or PDD, (3) participants aged between 0 and 16 years, (4) physical exercise intervention as the independent variable and (5) randomized and non-randomized controlled trials, cohort studies and pilot studies.

Studies were excluded due to the following criteria: letters, editorials, commentaries, unpublished manuscripts, dissertations, government reports, books and book chapters, conference proceedings, meeting abstracts, lectures and addresses, consensus development statements, guideline statements, systematic reviews, meta-analyses, or clinical practice guidelines.

**Information sources**

The search strategy for this review was developed in cooperation with a Health Sciences Librarian; a second librarian reviewed the search strategy for accuracy using the Peer Review of Electronic Search Strategies (PRESS) Checklist (McGowan et al., 2010). The following databases were searched for this review: Education Resources Information Center (ERIC), MEDLINE, ProQuest Nursing, PsycINFO, PubMed and SPORTDiscus. The search strategy was first developed in PubMed and then adapted to the other databases. A broad number of subject terms related to our topic were used in order to compensate for the differences in terminology that existed within each of the databases (Appendix 1). Searches were conducted from July to September 2014. The search results were exported to Endnote X6 and organized by database. All citations were then exported from EndNote X6 into Microsoft Excel for screening.

**Study selection**

For this review, we used a two-phase screening process to select eligible studies. Phase 1 involved the first and second authors independently reviewing the titles and abstracts of all retrieved articles for relevance. This phase involved articles being organized into one of three categories: relevant, possibly relevant and irrelevant. In phase 2, the same authors independently reviewed the articles and discussed disagreements with the aim of reaching a unanimous consensus for eligibility of the respective articles. Although not needed, a third reviewer was planned to be involved if agreement could not be reached.

**Assessment of risk bias**

The first and second authors each independently appraised the internal validity of the articles deemed relevant using the SIGN checklists (Harbour and Miller, 2001). Due to our inclusion of controlled trials and cohort studies, the SIGN checklist specific to the type of study appraised was used. The checklists were used to assess the level of selection, detection and confounding bias in the results of each study. An example of selection bias includes having significant differences between the experimental and control groups, where applicable, at the onset of the study. Examples of detection bias include a lack of clearly defined outcomes, lack of valid and reliable outcome measures and a lack of blinding of the outcome assessments. Finally, confounding bias occurs when potential confounders have not been considered or allowed for in the analysis (Harbour and Miller, 2001). A quantitative score or cut-off point was not used to determine internal validity of the studies; instead, the SIGN checklists helped the reviewers make a qualitative assessment regarding the internal validity of each study. A study was deemed to be of high overall quality if the majority of the SIGN criteria were met; there was little or no risk of bias and the results were unlikely to be changed by further research. An overall rating of acceptable was given when most criteria were met, if there was some risk of bias or if the conclusions could change with future research. Finally, a study was deemed to be of low overall quality when most criteria were not met or there were significant flaws in study design, meaning that conclusions would be likely to change with future research (Harbour and Miller, 2001). If consensus regarding internal validity of the articles was not reached through discussion between the first and second authors, a third reviewer was involved to help make the final decision. Only studies that were appraised as having adequate internal validity, and consequently low risk of bias, were included in our data extraction and synthesis of results.

**Data extraction and synthesis of results**

The second author extracted data from the scientifically admissible articles to build evidence tables in Microsoft
Word. Extracted items included the author and year; study design; composition of the sample including number of participants, their age and sex, and their diagnosis; the type of intervention and the frequency and dosage with which it was implemented; whether there was a control group and the activity in which the control group participated; the behavioural outcomes and how they were measured; the results of the behavioural outcomes; and whether there was a follow-up assessment. The first author independently reviewed the extracted data for accuracy. Due to the heterogeneity of the samples, interventions and outcomes in the admissible studies, a meta-analysis was not conducted. A qualitative synthesis of the results of the admissible studies was performed based on the guidelines for a best evidence synthesis (Slavin, 1995). Effect sizes (Cohen’s d) were calculated for the primary and secondary outcomes of studies that included a control condition. When mean values and standard deviations (SD) were reported, effect sizes were calculated using the following formula: 

$$d = \frac{\bar{x}_1 - \bar{x}_2}{SD_{pooled}}.$$

When only a t-statistic was reported, the following formula was used: 

$$d = t \sqrt{\frac{(n_1 + n_2)}{n_1 n_2} \left[ \frac{(n_1 + n_2)}{n_1 + n_2 - 2} \right]}.$$

The following thresholds were used to interpret the effect sizes: 0.2 was considered small, 0.5 was considered a medium effect and 0.8 was considered a large effect (Cohen 1992; Thalheimer and Cook, 2002).

**Reporting**

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) statement was used to organize and report the findings of this systematic review.

**Results**

**Study selection**

Our search retrieved 124 articles. We identified and removed 34 duplicates and screened 90 articles (titles and abstracts) for eligibility. Following this initial screening, 68 articles did not meet our selection criteria and were removed, and the remaining 22 articles underwent the full-text critical appraisal. The inter-rater agreement for the screening of articles was 93.8%. We accepted 13 articles as scientifically admissible for our review (Bahrami et al., 2012; Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Levinson and Reid, 1993; Movahedi et al., 2013; Nicholson et al., 2011; Oriel et al., 2011; Pan, 2010; Prupas and Reid, 2001; Rosenblatt et al., 2011; Rosenthal-Malek and Mitchell, 1997; Ward et al., 2013; Figure 1).

**Risk of bias within studies**

Inter-rater agreement of the critical appraisal was 90.9%, and an agreement on the inclusion of the two articles with discrepancies was reached following discussion with the independent reviewers. A total of 22 articles were critically appraised for inclusion in this review; 9 of these articles (40.9%) were deemed scientifically inadmissible because of poor internal validity (Celiberti et al., 1997; Chan et al., 2008; Kern et al., 1982, 1984; McGimsey and Favell, 1988; Magnusson et al., 2012; Morrison et al., 2011; Reid et al., 1988; Watters and Watters, 1980). One of the excluded studies was not relevant to our key questions as it did not include participants with ASD (McGimsey and Favell, 1988). The methodological weaknesses of the other eight excluded studies included selection bias (n = 8), detection bias (n = 8) and confounding bias (n = 8). An example of a common form of selection bias was that participants were selected for demonstrating high levels of stereotypic or maladaptive behaviours at the onset of the study, rather than selecting participants with varying levels of behaviours (e.g. Celiberti et al., 1997; Kern et al., 1982; Watters and Watters, 1980). An example of detection bias that occurred was that studies did not demonstrate the validity of their assessments and did not recognize that their outcomes may have been influenced by a lack of blinding (e.g. Morrison et al., 2011; Reid et al., 1988; Watters and Watters, 1980). Finally, an example of confounding bias that commonly occurred in the excluded studies was a lack of acknowledgement or accounting for the possibility that factors, such as additional therapy,
daily behaviour and familiarization with study personnel, may have accounted for some of the outcomes (e.g. Celiberti et al., 1997; Kern et al., 1982, 1984; Reid et al., 1988; Watters and Watters, 1980). The methodological quality of the included studies is presented in Table 1.

### Study characteristics

The study characteristics and demographic variables are presented in Table 2. There were five main types of exercise used in the 13 articles included in this review: jogging/running (5), horseback riding (4), martial arts (2), yoga and dance (1) and swimming (1). The majority of studies took place in a school setting (Bahrami et al., 2012; Levinson and Reid, 1993; Movahedi et al., 2013; Nicholson et al., 2011; Oriel et al., 2011; Prupas and Reid, 2001; Rosenthal-Malek and Mitchell, 1997), five were implemented in a community setting (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Pan, 2010; Ward et al., 2013) and the remaining study was set in a teaching hospital (Rosenblatt et al., 2011). Six of the studies included a control group, all of which comprised children with ASD (Bahrami et al., 2012; Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Movahedi et al., 2013; Pan, 2010), and three of which were randomized (Bahrami et al., 2012; Kern et al., 2011; Movahedi et al., 2013). Of these, three of the control groups participated in their regular treatment or activities while receiving no formal exercise (Bahrami et al., 2012; Movahedi et al., 2013; Pan, 2010) and the remaining three used a waitlist control design (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011).

Two of the studies reported the participants’ IQ scores (Gabriels et al., 2012; Rosenthal-Malek and Mitchell, 1997). Gabriels et al. (2012) had a clinical psychologist assess nonverbal IQ as part of their baseline screening and excluded participants with nonverbal IQ scores less than 40; participants included in the study had a mean nonverbal IQ of 95, ranging from 44 to 139. Participants in Rosenthal-Malek and Mitchell’s (1997) study had a mean IQ of 42.2, ranging from 36 to 48; however, it was unclear who provided the IQ assessment.

The majority of the interventions were implemented by trained professionals under the guidance of the authors (Bahrami et al., 2012; Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Movahedi et al., 2013; Rosenblatt et al., 2011; Ward et al., 2013), while the remaining six studies were implemented by the authors and trained research assistants. A high level of personnel support was provided for most of the studies as a participant to instructor ratio of at least 1:1 was maintained for 10 of the studies (Bahrami et al., 2012; Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Levinson and Reid, 1993; Movahedi et al., 2013; Nicholson et al., 2011; Prupas and Reid, 2001; Rosenblatt et al., 2011; Ward et al., 2013). Of the three remaining studies, one provided a 1:2 instructor-to-participant ratio (Pan, 2010) and the remaining two studies were unclear in their instructor-to-participant ratio (Oriel et al., 2011; Rosenthal-Malek and Mitchell, 1997).

There was a wide range in the frequency with which the intervention sessions were conducted. For example, four of the studies had participants engage in their intervention at a frequency of once per week (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Ward et al., 2013). One study ran their intervention two times per week (Pan, 2010), one was implemented three times per week (Nicholson et al., 2011) and two were implemented four times per week (Bahrami et al., 2012; Movahedi et al., 2013); the remaining studies had variable or unspecified

---

### Table 1. Quality of included studies (adapted from the SIGN checklists; Harbour and Miller, 2001).

<table>
<thead>
<tr>
<th>Author(s) (year)</th>
<th>Controlled for</th>
<th></th>
<th>Confounding</th>
<th>Overall study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selection bias</td>
<td>Detection bias</td>
<td>bias</td>
<td>quality</td>
</tr>
<tr>
<td>Bahrami et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>High</td>
</tr>
<tr>
<td>Bass et al. (2009)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>High</td>
</tr>
<tr>
<td>Gabriels et al. (2012)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Kern et al. (2011)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Levinson and Reid (1993)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Movahedi et al. (2013)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>High</td>
</tr>
<tr>
<td>Nicholson et al. (2011)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Oriel et al. (2011)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Pan (2010)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Prupas and Reid (2001)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Rosenblatt et al. (2011)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Rosenthal-Malek and Mitchell (1997)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Ward et al. (2013)</td>
<td>NA</td>
<td>✓</td>
<td>✓</td>
<td>Acceptable</td>
</tr>
</tbody>
</table>

SIGN: Scottish Intercollegiate Guidelines Network; NA: not applicable.
frequencies. The total dosage of exercise across all studies ranged from 80 min (in 10-min sessions; Prupas and Reid, 2001) to 56 h (in 1-h sessions; Bahrami et al., 2012; Movahedi et al., 2013; Table 2).

**Summary of evidence**

A summary of the study results are presented in Table 3. There were 11 different behavioural outcomes assessed in the 13 studies and of these, 3 primary categories of behaviour were addressed: stereotypic behaviour, cognition and attention (including on-task behaviour and academic responding) and social-emotional behaviour (including adaptive skills, social skills and problem behaviours).

**Stereotypic behaviour.** Six of the studies examined the impact of their exercise intervention on the stereotypic behaviour of their participants (Bahrami et al., 2012; Gabriels et al., 2012; Levinson and Reid, 1993; Oriel et al., 2011; Prupas and Reid, 2001; Rosenthal-Malek and Mitchell, 1997). Of these six studies, four intervened through short bursts of jogging and then measured stereotypic behaviour directly afterwards by recording the frequency of predetermined child-specific stereotypic behaviours from baseline recordings (Levinson and Reid, 1993; Oriel et al., 2011; Prupas and Reid, 2001; Rosenthal-Malek and Mitchell, 1997). One study used an experimental design to implement a martial arts intervention and assessed stereotypic behaviour through a standardized assessment tool (Bahrami et al., 2012), while the final study used another standardized tool and a waitlist control design to measure stereotypic behaviour before and after a 10-week horseback riding programme (Gabriels et al., 2012). Oriel et al. (2011) did not find a reduction in stereotypic behaviours following their jogging intervention; however, the other five studies found significant reductions.

Only two of the studies employed controlled trials to assess the impact of exercise on stereotypic behaviours (Bahrami et al., 2012; Gabriels et al., 2012). Bahrami et al. (2012) found statistically significant reductions in stereotypic behaviour following their intervention and no significant changes from post-intervention to the 30-day follow-up. Similarly, Gabriels et al. (2012) found statistically significant reductions in stereotypic behaviour following their horseback riding intervention, as well as statistically significant reductions in stereotypic behaviour from pre- to post-intervention in comparison to the waitlist control condition (Table 3). While effect sizes were only available for three of the studies that examined stereotypic behaviour (Bahrami et al., 2012; Gabriels et al., 2012; Rosenthal-Malek and Mitchell, 1997), the average effect was 1.6, which is large.

**Cognition and attention.** Three studies included measures of cognition or attention including on-task behaviour, academic responding and work performance. For example, Rosenthal-Malek and Mitchell (1997) found a significant increase in the frequency of correct academic responses given and a significant increase in the number of work tasks performed following 20-min bouts of jogging in five males aged 14–15 years with ASD. In a similar jogging intervention for 3- to 6-year-old children with ASD, Oriel et al. (2011) found a significant increase in the number of correct academic responses following a 15-min bout of exercise in comparison to when the participants did not engage in jogging. There was, however, no difference in the time spent engaged in academic tasks following the jogging intervention (Oriel et al., 2011). Finally, Nicholson et al. (2011) implemented a 15-min jogging intervention for four 9-year-old males with ASD. They found that the percentage of time spent engaged in academic tasks increased approximately 7.5% following the intervention period and then returned to near baseline levels at the 4-week follow-up (Nicholson et al., 2011). Only Rosenthal-Malek and Mitchell (1997) reported the necessary variables to calculate the effect of their intervention; however, the effect was large (Table 3).

**Social-emotional behaviour.** Children with ASD experience challenges with their social reciprocity, communication and behavioural skills (American Psychiatric Association, 2013). These difficulties can lead to the presence of difficult behaviours including aggression and non-compliance (Fox et al., 2002; Singh et al., 2006); however, they are also associated with challenges with social understanding, including empathy and joint attention (Baron-Cohen et al., 1985; Charman et al., 1997; Downs and Smith, 2004). Seven of the studies included in this review examined various aspects of social-emotional functioning in their participants following the exercise intervention (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Movahedi et al., 2013; Pan, 2010; Rosenblatt et al., 2011; Ward et al., 2013). Four of the these studies intervened with horseback riding (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Ward et al., 2013). Following their horseback riding intervention, Bass et al. (2009) found a statistically significant improvement in social responsiveness, measured by parent-report, in the experimental group in comparison to a waitlist control group who did not receive the intervention. Similarly, Ward et al. (2013) found significant improvements in their participants’ social interactions, although there was no control group. In contrast to the previous two studies, Kern et al. (2011) did not find any significant changes in parent-child interactions between the waitlist control and intervention periods. The final horseback riding intervention to include a measure of social-emotional functioning found significant improvements in adaptive behaviour, including communication, social and daily living skills, from pre- to post-intervention (Gabriels et al., 2012); however, these changes were not significant when compared to the waitlist control period (Gabriels et al., 2012).
Table 2. Study characteristics.

<table>
<thead>
<tr>
<th>Author(s) (year)</th>
<th>Study design</th>
<th>Sample (n, male%)</th>
<th>Diagnosis (criteria)</th>
<th>Age (years; range, M, SD)</th>
<th>Intervention (type, frequency and dosage)</th>
<th>Control (type, frequency and dosage)</th>
<th>Follow-up (time post-intervention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahrami et al. (2012)</td>
<td>Controlled trial</td>
<td>30, 87% male</td>
<td>ASD (DSM-IV-TR; formal diagnosis through institute for children with autism)</td>
<td>5–16 M=9.13 SD=3.27</td>
<td>Martial arts (n = 15): 1 session (60min), 4 days/week and 14 weeks (56h)</td>
<td>No formal exercise (n = 15): matched by age, gender and autism severity then randomized</td>
<td>30 days</td>
</tr>
<tr>
<td>Bass et al. (2009)</td>
<td>Controlled trial</td>
<td>34, 86% male</td>
<td>ASD (DSM-IV-TR; enrolled at autism treatment centre)</td>
<td>4–10*</td>
<td>Horseback riding (n = 19): 1 session/week, 1 h/session for 12 weeks (12h)</td>
<td>Waitlist (n = 15): regular treatment/activity during the 12-week intervention</td>
<td>NA</td>
</tr>
<tr>
<td>Gabriels et al. (2012)</td>
<td>Controlled trial</td>
<td>42, 86% male</td>
<td>Autistic or Asperger’s disorder (DSM-IV by licenced clinical psychologist)</td>
<td>6–16 M=8.7</td>
<td>Horseback riding (n = 42): 1 session/week, 1 h/session and 10 consecutive weeks (10h)</td>
<td>Waitlist (n = 16): regular treatment/activity for 10 weeks pre-intervention</td>
<td>NA</td>
</tr>
<tr>
<td>Kern et al. (2011)</td>
<td>Controlled trial</td>
<td>24, 75% male</td>
<td>ASD (family physician)</td>
<td>3–12 M=7.8 SD=2.9</td>
<td>Horseback riding (n = 24): 1 session/week, 1 h/session for 6 months (24h)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Levinson and Reid (1993)</td>
<td>Cohort</td>
<td>3, 67% male</td>
<td>Autism (DSM-III; confirmed by director of special education school)</td>
<td>11</td>
<td>Jogging (n = 3): 15-min session, walking on day 1 and jogging on day 2 for 5 weeks</td>
<td>Within-subject crossover: walking (n = 3); 15-min session, walking on day 1 and jogging on day 2 for 5 weeks</td>
<td>90 min</td>
</tr>
<tr>
<td>Movahedi et al. (2013)</td>
<td>Controlled trial</td>
<td>26, 85% male</td>
<td>ASD (DSM-IV-TR; formal diagnosis through institute for children with autism)</td>
<td>5–16 M=9.3 SD=3.38</td>
<td>Martial arts (n = 13): 1 session (60min), 4 days/week and 14 weeks (56h)</td>
<td>No formal exercise (n = 13): matched by age, gender and autism severity then randomized</td>
<td>30 days</td>
</tr>
<tr>
<td>Nicholson et al. (2011)</td>
<td>Cohort</td>
<td>4, 100% male</td>
<td>Autism (50%); Asperger’s disorder (50%); confirmed by school psychologist</td>
<td>9</td>
<td>Jogging (n = 4): 15-min jogging followed by 5-min walking cool down, 3 times/week for 2–5 weeks (multiple baseline design)</td>
<td>NA</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Oriel et al. (2011)</td>
<td>Cohort</td>
<td>9, 78% male</td>
<td>ASD (78%), ID (11%), DDelay (11%), all met criteria for ASD as defined by IDEA 2004</td>
<td>3–6 M=5.2</td>
<td>Jogging (n = 9): 15-min jogging followed by 15-min classroom task, 3 weeks in experimental condition (no. of trials/week unclear)</td>
<td>Within-subject crossover (n = 9); 15-min classroom task not preceded by jogging, 3 weeks in control condition (no. of trials unclear)</td>
<td>NA</td>
</tr>
<tr>
<td>Pan (2010)</td>
<td>Controlled trial</td>
<td>16</td>
<td>Autism (50%), Asperger’s syndrome (50%), DSM-IV (physician assessment)</td>
<td>6–9</td>
<td>Water exercise swimming programme (group A; n = 8): 90-min/session, 2 sessions/week for 10 weeks (30h)</td>
<td>Regular treatment/activity (group B; n = 8)</td>
<td>10 weeks for group A only</td>
</tr>
</tbody>
</table>
The other three studies to measure social-emotional functioning intervened through martial arts (Movahedi et al., 2013), swimming (Pan, 2010) and yoga and dance (Rosenblatt et al., 2011). Movahedi et al. (2013) found significant improvements in the experimental groups’ social interactions, in comparison to the control group, from pre- to post-intervention, and these changes were maintained at the 30-day follow-up. In comparison, Pan (2010) noted significant improvements in social competence, as well as decreased problem behaviours, from pre- to post-intervention; however, these changes were not significant when compared to the waitlist control group. Finally, Rosenblatt et al. (2011) found significant reductions in behavioural symptoms from pre- to post-yoga and dance intervention but did not include a control group in their study design.

Five of the seven studies that examined the impact of exercise on social-emotional behaviour reported the necessary variables to calculate an effect size. These effects ranged from negligible (0.02; Kern et al., 2011) to large (1.4; Movahedi et al., 2013), with an average effect of 0.7, which is considered to be large.

The exercise type. Five different types of exercise were used as a mode of intervention in the included studies: jogging, horseback riding, martial arts, yoga/dance and swimming. Of the five studies that intervened through jogging (Levinson and Reid, 1993; Nicholson et al., 2011; Oriel et al., 2011; Prupas and Reid, 2001; Rosenthal-Malek and Mitchell, 1997), two reported stereotypic behaviours as the sole outcome (Levinson and Reid, 1993; Prupas and Reid, 2001), two reported

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Study design</th>
<th>Sample (n, male%)</th>
<th>Diagnosis (criteria)</th>
<th>Age (years; range, M, SD)</th>
<th>Intervention (type, frequency and dosage)</th>
<th>Control (type, frequency and dosage)</th>
<th>Follow-up (time post-intervention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prupas and Reid</td>
<td>Cohort 4</td>
<td>4</td>
<td>PDD (50%), autism (25%), fragile X syndrome (25%), DSM-IV (school psychologist)</td>
<td>5–9</td>
<td>Multiple frequency walking and jogging (n=4): 10 min/session, 3 times/day (2 h apart) followed by 15 min of classroom observation. Implemented two times for each participant over a 2-week period (1 h)</td>
<td>Within-subject crossover (n=4): Single frequency walking and jogging 10 min/session, 1 time/day followed by 15 min of classroom observation. Implemented two times for each participant over a 2-week period</td>
<td>1 week</td>
</tr>
<tr>
<td>Rosenblatt et al.</td>
<td>Cohort 24, 92% male</td>
<td>ASD (physician referral)</td>
<td>3–16 M=8.9 SD=3.6</td>
<td>Yoga and dance: 8 sessions 45 min/session (frequency unknown; 6 h)</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Rosenthal-Malek and Mitchell (1997)</td>
<td>Cohort 5, 100% male</td>
<td>Autism (special education class enrolment)</td>
<td>14–15 M=14.88 SD=0.51</td>
<td>Jogging (n=5): 1–5 sessions/week, 20 min/session, for a total of 10 sessions = 200 min. Implemented in random order with control pre-condition</td>
<td>Within-subject crossover (n=5): regular academic tasks 1–5 sessions/week, 20 min/session, 10 sessions. Implemented in random order with experimental pre-condition</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Ward et al.</td>
<td>Cohort 21, 72% male</td>
<td>Autism (DSM-IV-TR; enrolment in school district’s autism programme)</td>
<td>5–10 M=8.1</td>
<td>Horseback riding: 1 session/week for the following schedule: 6 weeks on, 6 weeks off, 4 weeks on, 6 weeks off and 3 weeks on, 45–60 min/session</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

M: mean; SD: standard deviation; NA: not applicable; ASD: autism spectrum disorder; ID: intellectual disability; DD: developmental delay; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders – 4th ed.; DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders – 4th ed., text rev.; DSM-R-III: Diagnostic and Statistical Manual of Mental Disorders – revised 3rd ed.; IDEA 2004: Individuals with Disabilities Education Act 2004; PDD: pervasive developmental disorder; *M and SD are not reported in the table if they were not provided in the study itself.
Table 3. Summary of study results.

<table>
<thead>
<tr>
<th>Author(s) (year)</th>
<th>Outcomes of interest</th>
<th>Measures</th>
<th>Results</th>
<th>Effect size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahrami et al. (2012)</td>
<td>1. Stereotypic behaviour</td>
<td>1. Gilliam Autism Rating Scale – 2nd ed. (GARS-2)</td>
<td>1. Significant group by time interaction ($F(1, 28) = 22.09, p &lt; 0.001$) with the experimental group significantly decreasing their stereotypy scores by an average of 5.33 points from pre- to post-intervention ($t(14) = 5.94, p &lt; 0.001$). In contrast, the control group only decreased their scores by an average of 0.53 points pre- to post-intervention, which was not significant ($t(14) = 1.10, p = 0.29$). At the 30-day follow-up, there was a slight increase (0.87 points) in stereotypy scores for the experimental group, which was not significant ($t(14) = -0.91, p = 0.38$).</td>
<td>1. 0.9</td>
</tr>
<tr>
<td>Bass et al. (2009)</td>
<td>1. Social behaviour</td>
<td>1. Social Responsiveness Scale (SRS)</td>
<td>1. Significant group by time interaction on SRS overall score ($F(1, 20) = 4.92, p = 0.038$) with the experimental ($t(10) = 2.87, p = 0.017$), but not control ($t(10) = 0.108, p = 0.914$) group significantly improving pre- to post-intervention</td>
<td>1. 0.7</td>
</tr>
<tr>
<td></td>
<td>2. Sensory behaviour</td>
<td>2. Sensory Profile</td>
<td>2. Significant group by time interaction on overall Sensory Profile score ($F(1, 31) = 10.98, p = 0.002$) with the experimental ($t(18) = -7.29, p &lt; 0.01$), but not control ($t(13) = -1.77, p = 0.010$) group, significantly improving pre- to post-intervention</td>
<td>2. 0.5</td>
</tr>
<tr>
<td>Gabriels et al. (2012)</td>
<td>1. Self-regulation</td>
<td>1. Aberrant Behaviour Checklist – Community (ABC-C)</td>
<td>1. Significant decreases in ABC-C irritability ($p &lt; 0.001$), lethargy ($p &lt; 0.001$), stereotypic behaviour ($p &lt; 0.001$), hyperactivity ($p = 0.001$) and inappropriate speech ($p = 0.03$) from pre- to post-intervention. There were significant decreases in ABC-C irritability ($p = 0.004$), lethargy ($p = 0.007$), stereotypic behaviour ($p = 0.020$) and hyperactivity ($p = 0.008$) during the experimental condition in comparison to the waitlist control condition</td>
<td>1. 0.9</td>
</tr>
<tr>
<td></td>
<td>2. Adaptive skills</td>
<td>2. Vineland Adaptive Behaviour Scales – Interview Edition, Survey Form (VABS-2)</td>
<td>2. Significant improvements on the Adaptive Total Score ($p = 0.001$), Communication Raw Score ($p = 0.035$), Social Raw Score ($p = 0.016$), Daily Living Raw Score ($p = 0.011$) and Expressive Language ($p = 0.005$) domains of the VABS-2 from pre- to post-intervention. No significant changes in adaptive skills evident when comparing the waitlist control and experimental conditions</td>
<td>2. NA</td>
</tr>
<tr>
<td>Kern et al. (2011)</td>
<td>1. Parent-child</td>
<td>1. Timberlawn Parent-Child Interaction Scale</td>
<td>1. No significant changes in overall parent-child interactions when comparing the control to intervention period</td>
<td>1. 0.02</td>
</tr>
<tr>
<td></td>
<td>2. Sensory behaviour</td>
<td>2. Sensory Profile</td>
<td>2. No significant changes in overall sensory behaviour when comparing the control to intervention period</td>
<td>2. NA</td>
</tr>
<tr>
<td>Levinson and Reid (1993)</td>
<td>1. Stereotypic</td>
<td>1. Frequency of child-specific behaviours (including hand flapping, body rocking, intense staring, screaming, biting, etc.)</td>
<td>1. Prior to the control (walking) condition, the mean frequency of stereotypic behaviour was 73.4%. Post-control condition stereotypic behaviours increased to 75% and further increased to 89% at the 90-min follow-up. Prior to the experimental (jogging) condition, the mean frequency of stereotypic behaviour was 72%. Post-experimental condition stereotypic behaviours decreased to 54.5% and then increased to 82.8% at the 90-min follow-up</td>
<td>1. NA</td>
</tr>
<tr>
<td>Movahedi et al. (2013)</td>
<td>1. Social behaviour</td>
<td>1. Gilliam Autism Rating Scale – 2nd ed. (GARS-2)</td>
<td>1. Significant group by time interaction ($F(2, 48) = 14.91, p &lt; 0.001$) with the experimental group significantly decreasing their social interaction scores (indicating an improvement) by an average of 6.38 points from pre- to post-intervention ($t(12) = 6.17, p &lt; 0.001$). In contrast, the control group only decreased their scores by an average of 0.15 points pre- to post-intervention, which was not significant ($t(12) = 0.63, p = 0.55$). At the 30-day follow-up there was a slight increase (1.69 points) in social interactions scores (indicating greater dysfunction) for the experimental group, which was not significant ($t(12) = -1.65, p = 0.13$).</td>
<td>1. 1.4</td>
</tr>
<tr>
<td>Nicholson et al. (2011)</td>
<td>1. On-task behaviour</td>
<td>1. Behavioural Observation of Students in Schools (BOSS)</td>
<td>1. Mean time spent on in-task behaviour (academic engaged time) for the sample was 71.45% at baseline. This increased to 78.99% during the intervention and then returned to 73.63% at follow-up</td>
<td>1. NA</td>
</tr>
<tr>
<td>Oriel et al. (2011)</td>
<td>1. Stereotypic</td>
<td>1. Frequency of child-specific behaviours (including hand and arm flapping, body rocking and toe walking)</td>
<td>1. No significant differences in stereotypic behaviours were evident between the experimental and control conditions</td>
<td>1. NA</td>
</tr>
<tr>
<td></td>
<td>2. Academic responding</td>
<td>2. Frequency of correct/incorrect academic responses to a directive given by the teacher</td>
<td>2. Mean percentage of correct academic responding was 71.49% following the control condition and 82.57% following the experimental condition, which was a statistically significantly different ($p &lt; 0.05$)</td>
<td>2. NA</td>
</tr>
</tbody>
</table>

(Continued)
Table 3. (Continued)

<table>
<thead>
<tr>
<th>Author(s) (year)</th>
<th>Outcomes of interest</th>
<th>Measures</th>
<th>Results</th>
<th>Effect size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan (2010)</td>
<td>3. On-task behaviour</td>
<td>3. Percentage of time child is seated and consistently responding to teacher directives</td>
<td>3. No statistically significant differences in on-task behaviour were evident following the experimental (95.92% on task) and control (94.48% on task) conditions</td>
<td>3. NA</td>
</tr>
<tr>
<td>Malek and Mitchell (1997)</td>
<td>1. Stereotypic behaviour</td>
<td>1. Frequency of child-specific behaviours (including echolalia, body rocking, hand flapping, intense staring, etc.)</td>
<td>1. All participants had decreases in stereotypic behaviour following both the single frequency and multiple frequency conditions. Mean reduction for all subjects following the single frequency condition was 51.6%. Mean reduction for all subjects following the multiple frequency condition was 58.9%</td>
<td>1. NA</td>
</tr>
<tr>
<td>Reid (2001)</td>
<td>1. Stereotypic behaviour</td>
<td>1. Frequency of child-specific behaviours (including echolalia, body rocking, hand flapping, intense staring, etc.)</td>
<td>1. Significant reduction in Behavioural Symptom Index (p = 0.04) and atypicality (p = 0.02) from pre- to post-intervention for the entire sample. Significant reductions in Behavioural Symptom Index (p = 0.013), externalization (p = 0.04), internalization (p = 0.02), atypicality (p = 0.003) and depression (p = 0.02) for the participants aged between 5 and 12 years (n = 16) from pre- to post-intervention</td>
<td>1. NA</td>
</tr>
<tr>
<td>Rosenblatt et al. (2011)</td>
<td>1. General behaviour</td>
<td>1. Behavioural Assessment System for Children – 2nd ed. (BASC-2)</td>
<td>1. Significant improvement over the study period in sensory registration (F(5, 100) = 2.29, p &lt; 0.05), sensory sensitivity (F(5, 100) = 2.99, p &lt; 0.05) and reactions to auditory (F(5, 100) = 3.42, p = 0.007), visual (F(5, 100) = 3.50, p &lt; 0.05) and tactile (F(5, 100) = 3.82, p &lt; 0.05) sensory inputs</td>
<td>1. NA</td>
</tr>
</tbody>
</table>

NA: not applicable.

on both stereotypic behaviours and cognition and attention (Oriel et al., 2011; Rosenthal-Malek and Mitchell, 1997), while the remaining study reported outcomes relating solely to cognition and attention (Nicholson et al., 2011). Of these studies, effect sizes could only be calculated for Rosenthal-Malek and Mitchell’s (1997) intervention which found effect sizes of 3.0 and 2.0 for stereotypic behaviour and cognition and attention, respectively.

Four studies intervened through horseback riding (Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Ward et al., 2013), of which three examined the impact on social-emotional behaviour (Bass et al., 2009; Kern et al., 2011; Ward et al., 2013) and one examined the impact on both social-emotional and stereotypic behaviour (Gabriels et al., 2012). Effect sizes were not available for Ward et al. (2013) study, but the effect sizes of the remaining studies ranged from 0.02 to 0.9 with an average of 0.54, which is a moderate effect.

Two studies intervened through martial arts, one of which examined the impact on stereotypic behaviour...
only one study examined the impact of a yoga and dance intervention (Rosenblatt et al., 2011) and the behavioural outcome of interest was cognition and attention. However, effect sizes could not be calculated for this study. Similarly, only one study intervened through swimming (Pan, 2010) and then examined the impact on social-emotional behaviour. The effect size of this intervention was 0.6, indicating a moderate effect.

Generalization across the age span. Studies were split into three age bands according to the average age of the study participants. The age bands were early childhood (aged 0–5 years), middle childhood (aged 6–11 years) and adolescence (aged 12–16 years). Only one study was included in the early childhood age band (Oriel et al., 2011), which was a jogging intervention for which no effect size could be calculated. The majority (n = 11) of studies had an average age of participants in middle childhood and included all five types of exercise that have been discussed in this review. Effect sizes were calculated for six of these studies (Bahrami et al., 2012; Bass et al., 2009; Gabriels et al., 2012; Kern et al., 2011; Movahedi et al., 2013; Pan, 2010) and they ranged from 0.02 (Kern et al., 2011) to 1.4 (Movahedi et al., 2013).

Finally, only one study was included in the adolescent age band (Rosenthal-Malek and Mitchell, 1997). The intervention in question was jogging and had a large effect size of 3.0. None of the included studies controlled for age in their analyses.

Discussion

The purpose of this systematic review was to examine the impact of exercise interventions on the behavioural outcomes of children and youth aged ≤16 years. Following a comprehensive peer-reviewed search, two-stage screening process and critical appraisal, 13 articles were deemed scientifically admissible for inclusion. The results of this review are in agreement with previously published reviews on the impact of exercise interventions for individuals with ASD, all of which indicate a wide range of benefits (Lang et al., 2010; Petrus et al., 2008; Sorensen and Zarrett, 2014; Sowa and Meulenbroek, 2012). However, this review has built on the previous literature by providing an updated search of the empirical literature, using our aforementioned review methodology and by focusing solely on the behavioural benefits of exercise for children and youth.

The results of this review indicate that exercise can be an effective behavioural intervention for children and youth with ASD. However, due to the large variability in intervention frequency, intensity, type and dosage, the relative benefits of varying interventions are still somewhat unclear. However, it appears that the greatest benefits are a result of martial arts and horseback riding interventions, while the benefits of yoga/dance and swimming may be more limited. Jogging interventions can also be quite effective, although they have been limited by their methodological designs. There was large variability in the frequency with which the intervention sessions were implemented. For example, some studies intervened once per week, others consistently implemented as many as four sessions per week, while the frequency of some studies varied from week to week. Although virtually all studies resulted in improvements to the participants’ behaviour, the large variability in intervention frequency makes it challenging to determine an optimal frequency to produce the greatest improvements based on this review.

The impact of exercise intensity on behavioural outcomes is also not well understood in the literature. Levinson and Reid (1993) demonstrated that a 15-min jogging session resulted in greater reductions in stereotypic behaviour than a 15-min session of walking while also recording the exercise intensity pre- and post-intervention through heart rate taken via radial pulse. However, their analysis of the data did not account for changes in stereotypic behaviour relative to changes in heart rate (Levinson and Reid, 1993). Similarly, while Prupas and Reid (2001) aimed to have participants work at 65%–70% of their maximum heart rate, it was not directly measured or taken into account during analysis. It is possible that more intense bouts of exercise will result in greater behavioural improvements; however, the vast variability in activity type and lack of measure of intensity in the current literature make these associations unclear. It is recommended that future research include measures of exercise intensity, such as maximal and average heart rate during exercise, and employ monitors, such as accelerometers or heart rate monitors, during exercise. This would make objective comparisons of the benefits of different types of exercise more feasible, as well as help to determine the relationship between exercise intensity and behavioural outcomes in children and youth with ASD. Additionally, the use of activity and physiological monitors would help to establish fidelity to the exercise protocol (i.e. did they actually exercise for the prescribed amount of time and intensity), which can also impact the relative effectiveness of the exercise intervention.

Total intervention dosage is another important aspect to consider in the overall effectiveness of any intervention. It is likely that exercise interventions with a higher dosage and longer overall study duration, such as the 56-h martial arts (spread over four sessions/week for 14 weeks) implemented by Bahrami (2012) and Movahedi’s (2013) research groups, may result in greater and long-lasting improvements in behaviour. However, since no study has
investigated the impact of dosage on behavioural outcomes, it is difficult to establish an ideal dosage for optimal results. For instance, we do not know whether similar outcomes would be seen following 28 h, rather than 56 h, of martial arts training. Similarly, we do not know whether 5 min of jogging would result in equivalent reductions in stereotypic behaviour to those evident following 15–20 min. Although a ‘more is better’ approach is often taken with behavioural interventions, we simply do not know whether this holds true with regard to exercise and behavioural outcomes or whether there is a ceiling effect where benefits begin to plateau. It is also plausible that an optimal dosage could differ depending on the type of behaviour that is targeted for improvement. For example, exercise may only have an acute effect on stereotypic behaviour but could possibly have a more sustained impact on social-emotional behaviour. Future research should examine the relative impact of dosage on these different behavioural outcomes by including multiple measures throughout the study in order to determine whether a plateau in improvements is evident. Researchers could also randomize their groups to receive varying intervention dosages (while maintaining a similar exercise intensity) in order to further examine the impact of dosage on behavioural outcomes.

In addition to examining the impact of intervention dosage through multiple assessments, there is the need to conduct follow-up assessments in order to determine whether behavioural improvements are acute or sustained. Only six of the included studies had a follow-up assessment, all but one (Pan, 2010) of which were 1 month or less following the intervention. It is important that future research include longitudinal follow-up assessments, in addition to those immediately post-intervention, in order to understand whether exercise has a sustained effect on behavioural outcomes. Follow-up assessments can also be conducted more frequently for outcomes, such as stereotypic behaviour, that may only have short-term changes following exercise. Continuous and longitudinal follow-up assessments will enable us to understand how exercise effects behavioural outcomes and for how long, thus providing valuable information that can be used in creating community and home-based exercise programmes for children and youth with ASD.

The behavioural benefits evident in this review following exercise interventions may have more to do with the type of activity, rather than the intensity at which it was performed. This review identified a wide variety of types of exercise interventions, as well as various behavioural outcomes. For example, the exercises ranged from martial arts training to jogging to horseback riding; all of which may offer unique benefits. Although there was still a large amount of overlap in the behavioural benefits following these diverse exercise interventions, each of them likely offers unique physical experiences that could contribute to behaviour in different ways. Similarly, the structure imposed by the nature of the various interventions may also contribute to differences in behaviour. For example, the repetitious nature of martial arts training may be more beneficial for a child with ASD who thrives on structure, whereas a jogging intervention may not provide a strict enough routine for this same child.

In general, exercise appears to have the greatest benefit for children and youth with ASD by reducing stereotypic behaviours and improving social-emotional functioning. Stereotypic behaviours were reduced through jogging, horseback riding and martial arts interventions in four different studies. While social-emotional behaviour was improved through horseback riding, martial arts, yoga/dance and swimming interventions in six studies. In contrast, cognition and attention were only improved through jogging. It is possible that the group dynamic in a swimming, martial arts or horseback riding lesson may offer more opportunities for social interactions than a more individualized jogging programme. However, that is not to say that a jogging intervention could not improve social-emotional functioning or that swimming could not improve cognition and attention; there is simply no literature on these effects due to the narrow focus of most studies and the overall limited amount of research in this area. The high participant to instructor ratio in all of the included studies could also mean that improvements in social behaviour are a result of involvement in an intervention (and concurrent increase in social interactions) and not directly attributed to exercise. However, regardless of cause, any improvement in social-emotional behaviour may be beneficial for a child with ASD. Future research should include multiple types of exercise (randomized by group) within one study while controlling for exercise intensity, in order to further examine the impact of particular activities (e.g. jogging, swimming and horseback riding) on behavioural outcomes.

It is also plausible that many of the behavioural improvements evident in this review are related to improvements in executive functioning. Executive functioning refers to the family of neuropsychological control functions that are necessary for physical, cognitive and emotional self-control (Diamond, 2012; Pennington and Ozonoff, 1996). Previous research has suggested that executive dysfunction may be related to the repetitive and stereotypic behaviours frequently exhibited by individuals with ASD (Boyd et al., 2009; Kenworthy et al., 2009; South et al., 2007). There is currently no published research on the impact of exercise on the broader classification of executive functioning in children with ASD; however, research with children with typical development suggests that exercise may result in improvements to executive functioning (Best, 2010; Davis et al., 2011). Therefore, it is possible that some of the behavioural improvements evident in this review may be attributed to concurrent...
improvements in executive functioning warranting future research.

Executive functioning is highly interconnected with intellectual ability often resulting in positive relationships between the two variables (Duggan and García-Barrera, 2015). It has been estimated that approximately 40% of individuals with ASD have a comorbid diagnosis of intellectual disability (CDC, 2014), which may result in greater social and behavioural challenges (Matson and Shoemaker, 2009) and pose difficulties to full participation in intervention studies. Previous research has suggested that children with ASD who have higher IQs pre-intervention will have greater improvements post-intervention in comparison to their peers with lower IQs (Rogers, 1998; Rogers and Vismara, 2008). Children with lower intellectual abilities often pose greater challenges in terms of compliance, behaviour, and comprehension during intervention studies, and it is also possible that they may not respond as well to the intervention allowing for a bias towards research conducted on children with higher levels of functioning. Only two studies included in this review reported participants’ IQ (Gabriels et al., 2012; Rosenthal-Malek and Mitchell, 1997), one of which excluded participants with nonverbal IQ scores under 40 (Gabriels et al., 2012). It is unclear whether the remaining studies included participants with a range of intellectual functioning. Given the associations between ASD, intellectual disability, executive functioning, and behavioural outcomes, future research should include measures of intellectual functioning in order to control for it as a potential confounding factor of improvement in the other areas.

In addition to difficulties with comprehending intervention instructions, children with a comorbid diagnosis of intellectual disability may pose additional challenges in physically completing the exercise. Previous research has demonstrated that children with ASD have significant challenges with their motor skills and coordination (Liu and Breslin, 2013; Lloyd et al., 2013), and it is possible that these challenges can be attributed to, and potentially compounded by, an intellectual disability (Lahtinen et al., 2007; Vuijk et al., 2010; Wuang et al., 2008). As such, it may be beneficial for future research to assess and control for intellectual ability, as well as the participants’ physical ability to complete the exercise in order to account for these potential confounding factors.

Furthermore, as individuals with ASD are truly heterogeneous in their behaviour and overall functioning, it is important that interventions are implemented for individuals across the entire spectrum of functioning. Research conducted across the spectrum will help us to understand whether the impact of exercise interventions on the behavioural outcomes of children and youth with ASD varies with severity of autism symptoms and/or intellectual ability. This may have practical implications for implementing exercise interventions in a recreational setting for children and youth with ASD who may present with a range of unique abilities and challenges. It is important that exercise interventions are feasible in a community setting so that more children and youth with ASD can access and engage in exercise. However, given the high participant-to-instructor ratio in the studies included in this review, the feasibility outside of research is still questionable as many community-based programmes cannot offer one-to-one support. Future research should address this issue by conducting community-based research studies, as well as programme evaluations and effectiveness studies of exercise programmes that are currently being run by community and non-profit organizations (e.g. Special Olympics). Furthermore, the effectiveness of parent- and sibling-implemented home exercise programmes should also be investigated in order to increase the accessibility of exercise interventions for families who are unable to engage in research or attend community-based programmes while providing the one-to-one support evident in most studies.

Limitations

Limitations to this review stem from the lack of robust empirical research on exercise interventions for children and youth with ASD in the published literature. The wide range of exercise types, paired with the limited literature, makes it difficult to compare effectiveness between studies. Another inherent limitation that stems from a lack of robust research is that studies were excluded for methodological issues such as a high risk of bias or a lack of clear diagnosis of ASD. As a result, these exercise interventions were excluded from this review even though they may offer further insight into the effectiveness of exercise at improving the behavioural outcomes of children and youth with ASD. The exclusion of studies published in languages other than English limits our findings as we may be excluding relevant articles that could add to our synthesis of the literature. None of the studies reported effect sizes, and many studies did not even report the variables necessary to calculate effect sizes. This makes comparing the relative effect of the different interventions difficult, as we were only able to calculate effect sizes for approximately half of the studies (Table 3). We recommend that future intervention studies report the effect size of their outcomes in order to better understand the impact of the intervention. This review is further limited by the small sample sizes of many of the included studies, which means that these studies were not adequately powered to detect significant effects. However, this is common in behavioural intervention research for children with ASD, and the fact that many studies still had medium to large effects suggests the potential effectiveness of exercise interventions at improving behavioural outcomes in this population.

Finally, an important aspect of intervention effectiveness is whether the participants continued to exercise
following the completion of the programme, as maintaining the exercise may contribute to maintained behavioural improvements. However, none of the included studies examined whether the participants continued to exercise post-intervention. This limitation should be addressed in future intervention studies by including longitudinal follow-ups with measures of exercise maintenance and should include exercise maintenance as a covariate in analyses. Furthermore, exercise maintenance is important with regard to cost-efficiency because if it is effective at improving behavioural outcomes, and maintained over time, it may reduce the amount of intensive one-on-one behavioural therapy that is required. These limitations further emphasize the need for more studies with methodologically robust research designs on this topic.

**Conclusion**

This systematic review examined the impact of exercise interventions on various behavioural outcomes in children and youth aged ≤ 16 years with ASD. Results indicate that exercise may be an effective intervention for multiple behavioural outcomes in this population. More specifically, exercise may result in a decrease in stereotypic behaviours and improvements in social-emotional functioning, cognition and attention. This is important given the current need for evidence-based effective and sustainable interventions for children and youth with ASD.

Future research should continue to examine the impact of exercise on the behavioural outcomes of children and youth with ASD. Particular attention should be placed on the effectiveness of exercise interventions for younger children given the need for early intervention strategies for children with ASD. Furthermore, future research should be inclusive of the entire spectrum of children and youth with ASD by including baseline measures of IQ, objective ASD diagnosis and overall functioning, in order to determine whether exercise is more or less effective for particular sub-populations of children on the spectrum. There is also a need to include more longitudinal interventions and assessments in order to better assess whether exercise can result in both acute and sustained improvements in the behaviour of individuals with ASD. Finally, it may be beneficial for future researchers to include measures of exercise intensity, through activity and physiological monitoring, to examine the relative impact of exercise intensity on behavioural outcomes.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship and/or publication of this article.

**References**


Appendix 1

Boolean search terms and complete search phrase

Breakdown of Boolean search phrase:

1. Population
   a. Autism Spectrum Disorder
   b. Autistic disorder
   c. ASD
   d. Autis*
   e. Asperger Syndrome
   f. Pervasive developmental disorder
   g. PDD
   h. PDD-NOS

2. Intervention
   a. Physical Activit*
   b. Recreation
   c. Exercise
   d. Aerobic
   e. Cardio
   f. Cardiovascular
   g. Intervention*
   h. program

3. Demographic
   a. Child*
   b. Pediatric*
   c. Youth
   d. Adolescen*

4. Outcome
   a. Self-Stimulation
   b. On Task Behavio*
   c. Maladaptive Behaviour
   d. behavior*
   e. Aggression
   f. Challenging Behav*

Complete Boolean search phrase (developed in PubMed):

((((('Autism Spectrum Disorder' OR 'Autistic disorder' OR ASD OR Autis* OR 'Asperger Syndrome' OR 'Pervasive developmental disorder*' OR PDD OR PDD-NOS)) AND ('Physical Activit*' OR Recreation OR Exercise OR Aerobic OR Cardio OR cardiovascular)) AND ('Physical Activit*' OR Recreation OR Exercise OR Aerobic OR Cardio OR cardiovascular)) AND (Child* OR Pediatric* OR Youth OR Adolescen*)) AND Intervention* OR program) AND (‘Self-Stimulation’ OR ‘On Task Behavio*’ OR ‘Maladaptive Behaviour’ OR behavior* OR Aggression OR ‘Challenging Behav*’)}